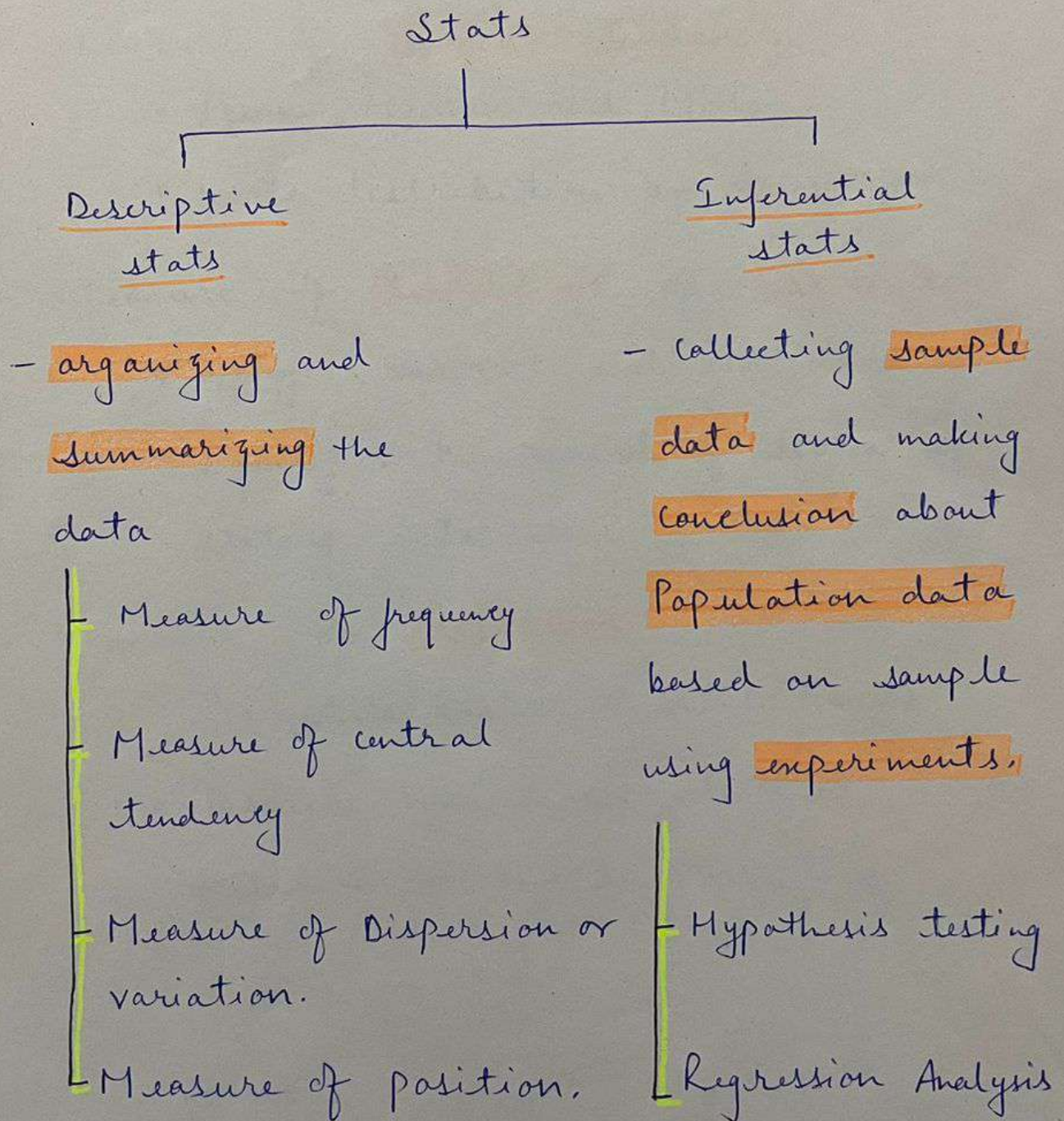


# Math Behind Data Science.

Statistic: It's the science of collecting, organizing and analyzing the data.

Data: It's fact or piece of information.





## Descriptive statistics:

It allows to characterize data based on its properties.

### — Measures of frequency:

- Count, Percent
- Shows how often something occurs

### — Measure of central tendency:

- Mean, Median and Mode.
- locate distribution by various points.

### — Measure of dispersion of variance:

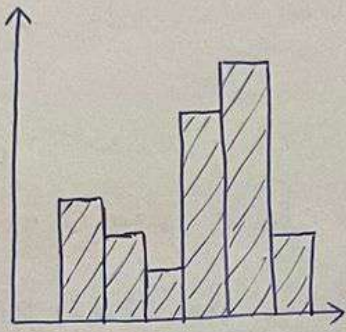
- Range, variance and std. deviation.
- Identifies the spread of scores by stating intervals.

### — Measure of position:

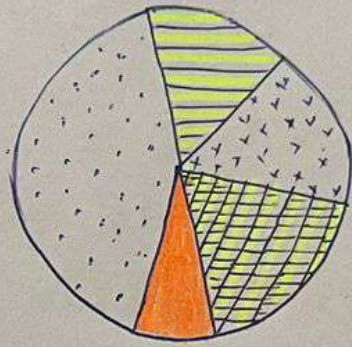
- Percentile Ranks, Quartile Ranks.
- Describe how scores fall in relation to one another.



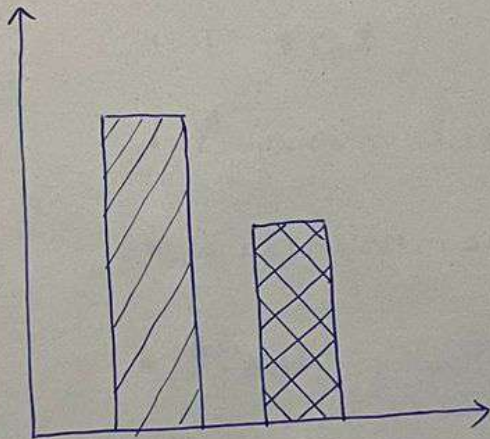
# Different types of plot used in descriptive stats:



Histogram



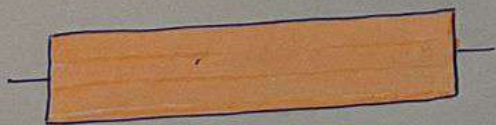
Pie chart



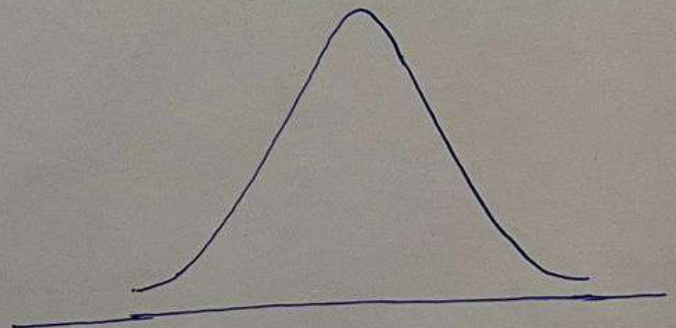
Bar chart



Candle chart.



Box plot



Distribution



## Inferential statistic:

Allows to make prediction of population data based on sample data.

Hypothesis testing: we use to test assumptions and draw conclusion about the population data based on available sample data.

- Z test
- F test
- T test
- Anova test
- Wilcoxon signed Rank test
- Mann whitney U test.

Regression analysis: Use to quantify how one variable will change with respect to other variable.

- Linear regression
- Nominal regression
- Logistic regression
- Ordinal regression.



## Sampling Technique:

**Sample**: Sample is a group of individuals who will actually participate in the research. selection of sample by different method for our observation is called **sampling method**. They are primarily of two types.

- **Probability sampling**: Involves **random selection**, allows to make strong statistical inferences about the whole group.
- **Non-Probability sampling**: Involves **non-random selection** based on **convenience** or other criteria, allows easy collection of data.

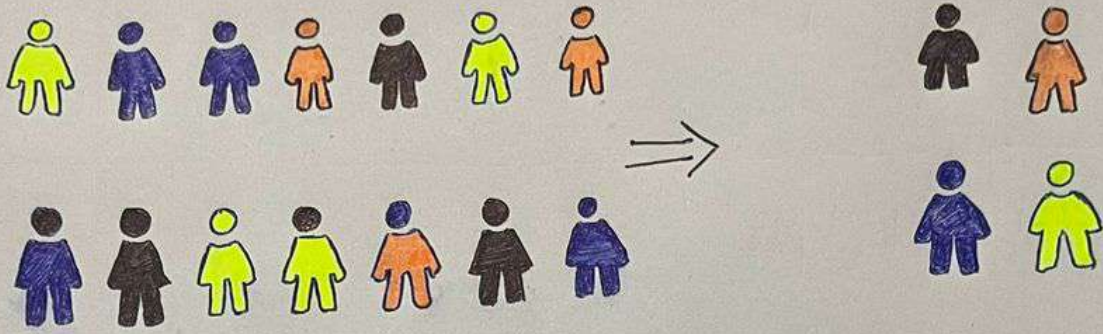
## Population vs Sample

**Population**: Entire group that we want to draw conclusions about.

**Sample**: Specific group of individual from which data will be collected.



## Population vs Sample.



Population.

Sample.

It is very important, your target population according to the purpose and practicalities of project.

If the population is very large, demonstrably mixed and geographically dispersed, it might be difficult to gain access to a representative sample. lack of sample affects validity of result.

N → Population size

n → Sample size.